Characterization and Recognition of Dynamic Textures based on 2D+T Curvelet Transform

Sloven Dubois, Renaud Péteri, Michel Ménard

SIViP manuscript No.
(will be inserted by the editor)

Characterization and Recognition of Dynamic Textures based on 2D+T Curvelet Transform

Sloven Dubois · Renaud Péteri · Michel Ménard

Received: date / Accepted: date

Abstract The research context of this article is the recognition and description of dynamic textures. In image processing, the wavelet transform has been successfully used for characterizing static textures. To our best knowledge, only two works are using spatio-temporal multiscale decomposition based on tensor product for dynamic texture recognition.

One contribution of this article is to analyse and compare the ability of the 2D+T curvelet transform, a geometric multiscale decomposition, for characterizing dynamic textures in image sequences. Two approaches using the 2D+T curvelet transform are presented and compared using three new large databases.

A second contribution is the construction of these three publicly available benchmarks of increasing complexity. Existing benchmarks are either too small, not available or not always constructed using a reference database.

Feature vectors used for recognition are described as well as their relevance, and performances of the different methods are discussed. Finally, future prospects are exposed.

Keywords Dynamic Textures · 2D+T Curvelet Transform · Spatio-temporal Multiscale Decompositions · Motion Recognition · Video indexing

Sloven Dubois
Université de Lyon, F-42023, CNRS, UMR5516, Laboratoire Hubert Curien, F-42000, Université de Saint-Étienne, Jean Monnet, F-42000, Saint-Étienne, France
Tel.: +33 477 915 797
Fax: +33 477 915 781
E-mail: sloven.dubois@univ-st-etienne.fr

Renaud Péteri
Laboratoire Mathématiques, Image et Applications, Avenue Michel Crépeau, 17042 La Rochelle, France
Tel.: +33 546 457 219
Fax: +33 546 458 240
E-mail: renaud.peteri@univ-lr.fr

Michel Ménard
Laboratoire Informatique, Image et Interaction, Avenue Michel Crépeau, 17042 La Rochelle, France
Tel.: +33 546 458 296
Fax: +33 546 458 242
E-mail: michel.menard@univ-lr.fr

1 Introduction

1.1 Context

Our visual world is composed of many complex structures and motions. Our human biological visual system has the potentiality to acquire, integrate, and interpret all of this complex information and to provide the ability to navigate easily through it. When looking at a scene, our brain instantly recognizes and characterizes regions of different appearances and motions.

The computer vision community is involved in studying and mimicking the human visual system and its ability to see and interpret our world. When a computer vision system acquires a natural scene, it should also be able to segment, characterize and interpret the different regions contained in the image, such as forest, lake, river, mountains, sky, etc.

In many situations, huge portion of our visual world is perceived as texture. Thus texture has become, in recent years, a fundamental characteristics for describing the image content (for example MPEG-7 descriptors [33]). The extension of these visual features to the temporal dimension leads to some new challenges. The notion of texture in image sequences raises many questions: what are textures in videos? To what extent 2D+T textures are different from static ones, or are they simple extension to 3D of 2D structures? What

are the phenomena leading to 2D+T textures?

Some answers are given in a pioneer work by Nelson and Polana [20,27]. The authors categorize events occurring in an image sequence into three classes: (1) spatially periodic pattern with temporal periodic motion, (2) spatially bounded shape with temporal periodic motion and (3) spatially bounded shape without temporal periodic motion. The first class is called *Dynamic Textures*, or more rarely *Temporal Textures*.

Defining formally what is a spatial texture is a difficult and hazardous task. Adding the temporal dimension further complicates its definition. According to Yves Meyer in the Workshop "An interdisciplinary approach to Textures and Natural Images Processing" [19], textures are "a subtle balance between repetition and innovation". Given a proper definition of dynamic textures is a notoriously difficult problem. They can not only be considered as a simple extension of static textures to the time domain, but as a more complex phenomenon resulting from several dynamics. It is possible to define briefly a dynamic texture as a time varying phenomenon with a certain repetitiveness in both space and time. In [13,25], we define a dynamic texture more precisely as follows:

A natural, artificial or synthetic image sequence may contain a static texture component and/or a dynamic texture component. This latest one is composed of at least one dynamic texture. A dynamic texture is a textured pattern that can be rigid or deformable. This pattern has a motion induced by a force which can be internal, external or created by camera motions. This motion can be deterministic or stochastic. Dynamic textures are composed of modes, which may overlap, characterized by repetitive spatial and temporal phenomena.

A flag flapping in the wind, fire, smoke, ripples at the surface of water, waving trees, traffic, an escalator, etc., are all examples of dynamic textures. Two examples are shown in Figure 1 (a flapping flag and grass). Each image sequence is viewed as a 3D data cube where cuts make it possible to observe dynamic texture motions.

For more details about this definition or dynamic textures notion, one can refer to [13,25].

After the pioneer work of Nelson and Polana [20, 27], the number of major publications on the topic of dynamic textures has risen sharply. This growing interest can be explained by a large field of applications: videos synthesis [10,6] (realistic dynamic texture synthesis for animations, video games, video inpainting), spatio-temporal segmentation [17] (to detect a perturbation in a given dynamic texture, to build video summaries), video surveillance [26] (to detect an accident in traffic, to detect forest fires, to characterize and supervise the motion of a crowd), video indexing [29,12] (to perform elaborate semantic queries), dynamic background subtraction [7,1], tracking [22] (to follow and to analyze the evolution of given phenomena), etc.

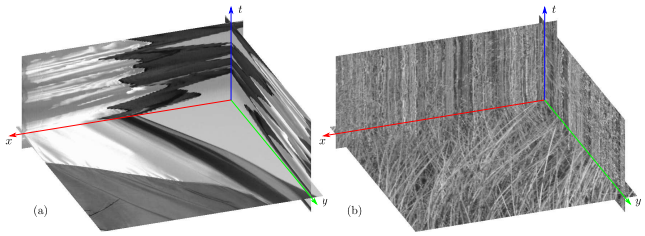

Fig. 1 *2D+T sections of two dynamic textures: (a) a flag flapping in the wind, (b) grass. Here, a dynamic texture is seen as a 3D data cube cut at voxel $O(x,y,t)$, giving three planes $(\vec{x}O\vec{y})$, $(\vec{x}O\vec{t})$ and $(\vec{y}O\vec{t})$.*

Our research context is the recognition and description of dynamic textures [24,37]. Our main goal is to obtain representative features of dynamic textures, *i.e.* the most compact and discriminative as possible. For a brief survey on this topic, one could refer to [8].

1.2 Outline of the article

In our context of dynamic texture characterization, previous works can be classified according to the following taxonomy: methods based on optical flow [20,28, 15], methods based on spatio-temporal filtering [31], methods computing geometric properties in the spatio-temporal volume [21,38], methods based on Linear Dynamical Systems [10,6] and methods using spatio-temporal transforms [29,12].

A natural tool for multiscale analysis is the wavelet transform. In the field of image processing, the wavelet transform has been successfully used for characterizing static textures. For instance, Gabor wavelets have been used for computing the texture descriptor of the MPEG-7 standard [33]. A natural idea is to extend these multiscale decompositions to the time domain in order to characterize dynamic textures.

To our knowledge, only two works have been using the spatio-temporal multiscale decompositions for characterizing dynamic textures. In 2002, J.R. Smith *et al.* [29] are proposing one spatio-temporal wavelet decomposition and analyze the impact of one feature descriptor on a small unavailable database. In 2009 [12], we have proposed three other spatio-temporal multiscale transforms for analyzing dynamic textures. This three methods are compared with the one of J.R. Smith *et al.* on a more complex unavailable database using one feature vector.

These two works are using spatio-temporal wavelet transforms based on tensor product of 1D transforms. However, as it will be mentioned in the next section, these multiscale transforms fail to represent and detect more geometric signals (lines, curves, ...). This drawback has been overcome by the emergence of several multiscale geometric transforms.

The major contribution of this article is to use the 2D+T curvelet transform (presented in Section 2) for extracting descriptors to the purpose of dynamic texture recognition. After briefly presenting the theory of this geometrical multiscale decomposition (Section 2.1), we describe the use this transform directly or within a formal model (Section 2.2). The formal model permits to decompose a dynamic texture into different components [13]. To our knowledge, it is the first time that dynamic texture recognition is performed from the components of the decomposition model.

Another contribution of this article is the construction of three new large datasets (presented in Section 3.2) available on the DynTex database website [25] for relevant testing and comparison with other approaches. In previous papers [29,12], the authors were using non available small database. This a limitation when it comes to study relevance of features and to compare different recognition multiscale approaches.

In Section 3.3, a comparison between the different multiscale approaches (approaches based on curvelet transform versus wavelet decompositions using filter product) is performed. Finally, obtained results are discussed and prospects are exposed in Section 3.4.

2 The 2D+T Curvelet Transform for dynamic texture analysis

It is undeniable that the wavelet transform has had a major impact in many applications of signal and image processing. However, for 2D signal, it fails in representing and detecting objects composed of anisotropic elements, such as lines or curves. For this reason, recent years have seen the emergence of several multiscale geometric transforms: the bandelet transform [18], the ridgelet transform [3], the curvelet transform [9], etc..

The curvelet transform has been designed for improving the limitations of the wavelet transform: while wavelets catch 1D singularities, curvelets can detect structures of higher dimensional structures (of co-dimension 1, *i.e.* curves in images).

The curvelet transform has been recently extended to the third dimension [36,5] where it is sparse for representing smooth surfaces.

In Section 2.1, the curvelet transform theory is presented. We detail why this decomposition is optimally sparse for the representation of dynamic texture wavefronts. After, in Section 2.2 a formal model is recalled and its relevance for characterizing dynamic textures is discussed.

2.1 2D+T Curvelet transform

Similarly to the wavelet decomposition construction, the 3D curvelet transform is the projection of a function $f \in \mathcal{L}^2(\mathbb{R}^3)$ on a basis of functions $\mathcal{L}^2(\mathbb{R}^3)$. A collection of coefficients $c(j,\ell,\mathbf{k})$ is obtained as the scalar product of $\mathcal{L}^2(\mathbb{R}^3)$ between the function f and the curvelet analysis functions $\varphi_{j,\ell,\mathbf{k}}$, also called atoms:

$$c(j,\ell,\mathbf{k}) := \langle f, \varphi_{j,\ell,\mathbf{k}} \rangle = \int_{\mathbb{R}^3} f(\mathbf{x}) \overline{\varphi_{j,\ell,\mathbf{k}}(\mathbf{x})} d\mathbf{x} \quad (1)$$

where $\mathbf{x} = (x,y,z)^T$ represents the coordinates of a voxel in the 3D cube, $\varphi_{j,\ell,\mathbf{k}}$ is the curvelet atom at scale $j \in \mathbb{Z}$, in direction $\ell \in \mathbb{Z}$ and position $\mathbf{k} = (k_1,k_2,k_3)$. Atoms $\varphi_{j,\ell,\mathbf{k}}$ are built by the composition of a translation $\mathbf{x}_{\mathbf{k}}^{(j,\ell)}$ and of a rotation R_ℓ of the atom φ_j:

$$\varphi_{j,\ell,\mathbf{k}} = \varphi_j \left(R_\ell \left(\mathbf{x} - \mathbf{x}_{\mathbf{k}}^{(j,\ell)} \right) \right) \quad (2)$$

The mother curvelet atom φ_j is expressed in the frequency domain by the mean of the Fourier transform, $\widehat{\varphi}_j(\boldsymbol{\omega}) = U_j(\boldsymbol{\omega})$, that can be written in polar coordinates as:

$$U_j(r,\theta) = 2^{-3j/4} W_j\left(2^{-j} r\right) V_j\left(\frac{2^{\lfloor j/2 \rfloor}\theta}{2\pi}\right) \quad (3)$$

The support of $U_j \in \mathbb{C}$ is a polar wedge (see Fig. 2.(c)) defined by the support of $W_j \in \mathbb{C}$ and $V_j \in \mathbb{C}$, representing respectively a radial window (see Fig. 2.(a)) and an angular window (see Fig. 2.(b)).

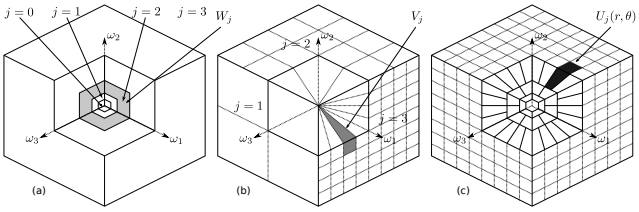

Fig. 2 *Discrete frequency tiling. The light and the mean gray colors represent respectively the window $V_{j,\ell}(\boldsymbol{\omega})$ and $W_j(\boldsymbol{\omega})$. The composition of two windows $U_{j,\ell}(\boldsymbol{\omega})$ is colored in black.*

For more information on the 3D discrete curvelet transform, one can refer to [36,5].

Transition from 3D to 2D+T is not trivial. Indeed, in the 3D case, distances between a pixel center and its

6-connex neighbors are the same ($\Delta_x = \Delta_y = \Delta_z$). In the 2D+T case, the spatial distance between two pixels is different from the distance along the time axis ($\Delta_x = \Delta_y \neq \Delta_t$). There is a relationship that can be written as $\Delta_z = \alpha.\Delta_t$ with α a constant that enables to keep the homogeneity between spatial and temporal variables. The constant α is homogeneous to speed and can be adapted to the considered video.

It has been shown in [11,4,13] that the 2D+T Curvelet Decomposition is relevant for extracting non-local phenomena propagating temporally. As detailed in the next section, a dynamic texture is indeed composed of different modes, constituted by wavefronts and local phenomena. The Figure 4.(a) shows that the spatio-temporal edges, created by the wavefront component, are well defined. The 2D+T Curvelet Decomposition thus seems particularly interesting to model these long range waves. Using energies of the 2D+T Curvelet Decomposition, we have shown that it is possible to spatio-temporally segment dynamic textures occurring in a video. In fact, the more a wavefront with a given frequency, orientation ℓ and scale j is important in a video, the higher the energy of the corresponding curvelet will be. In the frequency domain, a dynamic texture wavefront generates a high response to the $\varphi_{j,\ell,\mathbf{k}}$ atom. This observation led us to construct a new spatio-temporal segmentation algorithm based on an octree structure using as homogeneity criterion the energies of the coefficients of the 2D+T curvelet transform. A result obtained with this algorithm is presented in Figure 3. For other results and for more informations on the segmentation process, one can refer to [11]. This spatio-temporal segmentation algorithm shows that the coefficients of the 2D+T curvelet transform contain a discriminative information. This one can be later used for recognizing different dynamic textures.

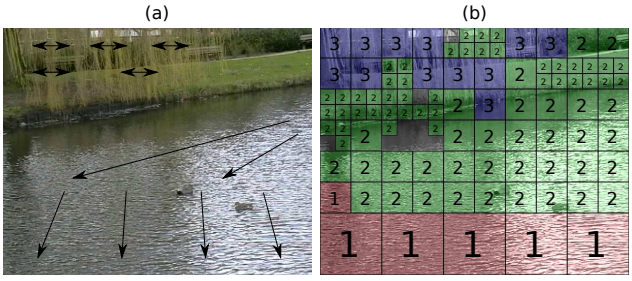

Fig. 3 *(a) Original video. Main spatio-temporal directions are symbolized by arrows. (b) Segmentation results of video using the energies of the coefficients of the 2D+T curvelet decomposition. Each color (red, green and blue, identified respectively by 1, 2 and 3) represents a distinct area. A non colored area (or non labeled area) corresponds to an ambiguity region.*

2.2 Dynamic texture decomposition based on a formal model

As mentioned previously, a dynamic texture is often described as a time varying phenomenon with a certain repetitivity in both space and time. Many dynamic textures are composed of visually relevant modes. For instance on Figure 4.a showing an image sequence of sea waves, two modes can be observed: the high-frequency motion of small waves (cf. Figure 4.a.2), carried by the overall motion of the internal wave (cf. Figure 4.a.1). The process gets more complex when the two phenomena overlap with each other (cf. Figure 4.a.3). These two modes can also be observed on the image sequence of waving trees on Figure 4.b.

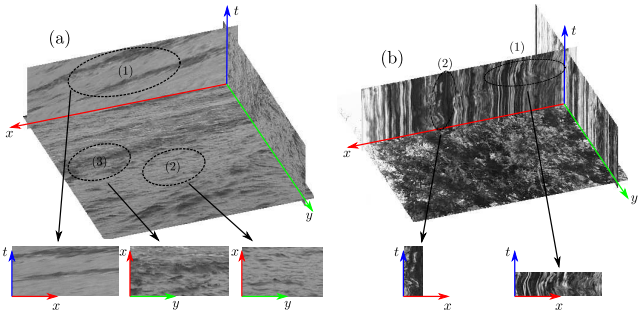

Fig. 4 *2D+T slices of two dynamic textures. One can observe several wavefronts (1), local oscillating phenomena (2) and a mixture of both of them (3).*

Following the above observations, as well as different works on video synthesis [16] and the study of the DynTex database [25], we have introduced in [13] a formal model for several kinds of dynamic textures. For a self-contained paper and better understanding, this model is summarized here. A dynamic texture Υ_i can be modeled as the superposition of large scale wavefronts and local oscillating phenomena. It can thus be defined as:

$$\forall i, \Upsilon_i^{\Omega_i}(\mathbf{x}) = \mathcal{P}_i(\mathbf{x}) + \mathcal{L}_i(\mathbf{x}) \quad (4)$$

where Ω_i represents the spatio-temporal support of dynamic texture Υ_i, $\mathcal{P}_i$ and $\mathcal{L}_i$ are two functions describing respectively the wavefront and local phenomena composing a dynamic texture Υ_i. The carrying wave $\mathcal{P}_i$ is the most complex phenomenon, and depends on the considered image sequence. It is characterized by its propagating speed, its direction and its degree of stationarity. Functions $\mathcal{P}_i$ propagate texture information given by local oscillating phenomena. Local phenomena $\mathcal{L}_i$ differ from the carrying wave by being purely local. For more information and description of this model, one

can refer to [13].

Model (4) is well adapted for the following dynamic textures:

- deformable textured patterns with stochastic or deterministic motion, such as fluid flows (lake, sea, water stream, *etc*), oscillations generated by wind (grass, trees, flag, *etc*), smoke propagation, *etc.*
- rigid textured patterns with deterministic motion such as an escalator, a windmill, *etc.*
- discrete textures with stochastic motion such as fish shoal, insect swarm, *etc.*

Analyzing dynamic textures represented by this model, results in decomposing them into local oscillating phenomena and non local wavefronts. Recent works for decomposing images and videos [2,30] seem relevant for extracting these components. The Morphological Component Analysis has been chosen because of the richness and the flexibility of the available dictionary, which is crucial considering the complexity of dynamic textures. For a complete description of the Morphological Component Analysis framework, one can refer to [30].

A crucial point in the Morphological Component Analysis approach is the definition of the dictionary. An unsuitable choice of transformations will lead to non sparse and irrelevant decompositions of the different dynamical phenomena present in the sequence. It is therefore necessary to associate each component of our model with the most representative bases.

We have shown that the 2D+T Curvelet Transform is relevant for extracting non-local phenomena propagating temporally. It thus seems particularly interesting to model long range waves present in a dynamic texture. The second part of the model is composed of locally oscillating phenomena that can be extracted using the 2D+T Local Cosine Transform. The Morphological Component Analysis dictionary is hence composed of these two bases, and enables to obtain results presented on figure 5.

The original version of the Morphological Component Analysis algorithm applied on an image sequence is very consuming in matter of computation time. Indeed, for one typical dynamic texture from the DynTex database (125 video frames of spatial resolution 720×576 pixels), the decomposition takes approximately 10 hours using a fairly powerful computer[1]. With the new adaptive thresholding strategy that we introduced in [11], this computation time is reduced to less than 2 hours.

The obtained components using the Morphological Component Analysis algorithm can be used for extracting characteristic features: some related to the geometry of the dynamic texture (main motion direction, uniformity of the overall movement, etc.) and some characterizing more local phenomena (speed, local vortex, etc.). To our knowledge, it is the first time that components of different dynamic content are used for video recognition. Feature vectors based on this decomposition are computed and tested in the next section.

[1] 64-bit processor, 3.2GHz with 26Go Ram

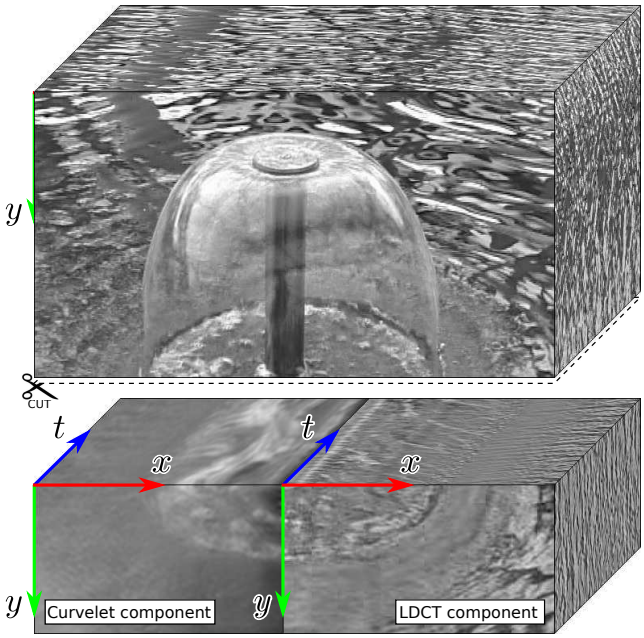

Fig. 5 *Decomposition results of a dynamic texture using the Morphological Component Analysis. The carrying wave is retrieved using the 2D+T Curvelet Transform while the local phenomena is obtained by the 2D+T Local Cosine Transform.*

3 Indexation of Dynamic Textures

One major contribution of this article concerns recognition of dynamic textures with multiscale methods. Different approaches are compared: two methods based on the 2D+T curvelet decomposition (see previous section) and four spatio-temporal multiscale decompositions based on tensor product [29,12]. These six approaches are tested with three new available databases. The main objective is to evaluate these approaches and identify the most relevant descriptors.

3.1 Experimental protocol

Each of the experiments have been set as follows: (1) analysis of image sequences using a spatio-temporal decomposition, (2) computation of descriptors and con-

struction of a feature vector, and (3) leave-one-out cross-validation for studying each feature relevance.

(1) Spatio-temporal analysis

Our descriptor vectors are constructed from different multiscale transforms:

- Spatial Wavelet Decomposition [12]: this approach uses the wavelet decomposition image per image. In this case, there is no consideration on the temporal correlation between two successive frames. For each image and for each scale of multiresolution analysis, the approximation sub-band and three details subbands are computed.
- Temporal Wavelet Decomposition [12]: the first method considers a video frame per frame, and is thus a purely spatial method. The second natural approach is to perform the multiresolution analysis in the time direction. For each pixel of a dynamic texture video, the temporal profile is extracted and its one dimensional wavelet transform is performed.
- 2D+T Wavelet Decomposition [12]: whereas the first method is a purely spatial decomposition and the second one is a temporal decomposition, the third method performs decomposition spatially and temporally. This extension to the temporal domain of the 2D discrete wavelet transform is done using separable filter banks. As in the 2D case, a separable 3 dimensional convolution can be factored into one-dimensional convolution along rows, columns and image indexes of the video. For a given video, seven detail subbands and one approximation subband are computed for each scale.
- J.R. Smith *et al.* Wavelet Decomposition [29]: this transform is similar to the 2D+T Wavelet Decomposition, except that the temporal filter is applied two times at each resolution step so that the video is decimated twice spatially and twice temporally. The authors obtain, for one video, fifteen detail subbands and one approximation subband.
- 2D+T Curvelet Decomposition (*cf.* Section 2.1).
- Components from the Morphological Component Analysis decomposition using the 2D+T Curvelet Transform and the 2D+T Local Cosine Transform (*cf.* Section 2.2).

Each method is identified by an index: $m = \{fpf, t, xyt, xy2t, curv, mca\}$ (indexes are following the order of the list above).

(2) Construction of feature vectors

The usual way to characterize 2D texture using wavelet decomposition is to build feature vectors from detail subbands [33]. The following descriptors are computed:

- average of detail subbands $\mu_m^{(j,\ell)}$,
- standard deviation of detail subbands $\sigma_m^{(j,\ell)}$,
- mean energy of detail subbands $E_m^{(j,\ell)}$,
- entropy of detail subbands $H_m^{(j,\ell)}$.

In the case of the 2D+T Local Cosine Transform used in the Morphological Component Analysis framework, detail subbands do not exist. As illustrated in Figure 6, the coefficients of this transform are divided into several subbands. Each subband corresponds to a set of oriented frequencies of similar scales. Descriptors are computed in the same way than for multiscale decompositions.

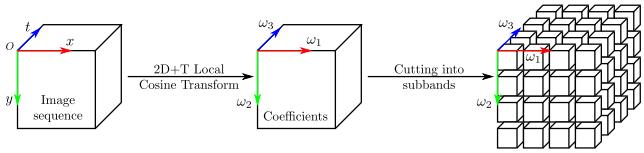

Fig. 6 *Coefficients splitting of the 2D+T Local Cosine Transform for obtaining detail subbands similarly to wavelet decomposition.*

In our aim of indexing dynamic textures, five different features are built (four are constructed directly from one descriptor and one is constructed from the concatenation of these four descriptors):

- Feature vector based on the average of detail subbands:

$$\mathbf{S}_m^{\mu} = \left(\mu_m^{(1,1)}, \cdots, \mu_m^{(1,N_\ell^1)}, \cdots, \mu_m^{(N_j,1)}, \cdots, \mu_m^{(N_j,N_\ell^j)} \right) \tag{5}$$

  with N_ℓ^j being the number of orientations at scale j and N_j the number of scales.
- Feature vector based on the standard deviation of detail subbands:

$$\mathbf{S}_m^{\sigma} = \left(\sigma_m^{(1,1)}, \cdots, \sigma_m^{(1,N_\ell^1)}, \cdots, \sigma_m^{(N_j,1)}, \cdots, \sigma_m^{(N_j,N_\ell^j)} \right) \tag{6}$$

- Feature vector based on the energy of detail subbands:

$$\mathbf{S}_m^{E} = \left(E_m^{(1,1)}, \cdots, E_m^{(1,N_\ell^1)}, \cdots, E_m^{(N_j,1)}, \cdots, E_m^{(N_j,N_\ell^j)} \right) \tag{7}$$

– Feature vector based on the entropy of detail subbands:

$$\mathbf{S}_m^H = \left(H_m^{(1,1)}, \cdots, H_m^{(1,N_\ell^1)}, \cdots, H_m^{(N_j,1)}, \cdots, H_m^{(N_j,N_\ell^j)} \right) \tag{8}$$

– Feature vector based on the previous characteristics of detail subbands:

$$\mathbf{S}_m^A = \left(\mathbf{S}_m^\mu, \mathbf{S}_m^\sigma, \mathbf{S}_m^E, \mathbf{S}_m^H \right) \tag{9}$$

For a given video database, a set of features $\mathbf{S}_{m,c,i}^d$ is obtained, with m representing the spatio-temporal analysis method, $d = \{\mu, \sigma, E, H, A\}$ is the descriptor used and i the i-th sample of class c of the base.

The feature vector is normalized as follows:

$$\forall n, \forall r, \forall j, \quad \mathbf{S}_{m,r,j}^d(n) = \frac{\mathbf{S}_{m,r,j}^d(n) - \min\limits_{c,i} \mathbf{S}_{m,c,i}^d(n)}{\max\limits_{c,i} \mathbf{S}_{m,c,i}^d(n) - \min\limits_{c,i} \mathbf{S}_{m,c,i}^d(n)} \tag{10}$$

with n representing the parameter index in the feature vector.

Experiments were also conducted with other parameters (variation of the number of scales, different normalization of feature vectors, etc.). The best results are presented in the next section.

(3) Leave-one-out cross-validation

For each approach m and each feature vector $\mathbf{S}^d$, its relevance is studied by computing a recognition rate from a confusion matrix using the leave-one-out cross-validation. This method is used to estimate how accurately a model will perform in practice [32]. In our case, it is possible to use it to study the relevance of our feature vectors and therefore our wavelet-based approaches. This cross validation method works well for small datasets. Indeed for a small number of samples, the within cluster variance can evolve quickly. The procedure is as follows:

– For each class c, its center is computed.
– For each element p of each class c:
  – Compute the center of class c without element p.
  – Find the nearest class r of element p.
  – If r is different from the original class of element p, a misclassification is recorded.

The leave-one-out cross-validation leads to a confusion matrix representative of the feature vector relevance. The diagonal elements of this matrix are the well classified samples and allows the computation of a recognition rate.

The Euclidian metric is currently used in the leave-one-out cross validation for computing the distance between two samples. It is also possible to use another metric, for instance the Mahalanobis distance [35]. It differs from Euclidean distance in that it takes into account the correlations of the data set and is scale-invariant.

3.2 Databases used

Each multiscale decomposition is tested for all descriptors (described in the previous section) on three new databases of increasing complexity. Image sequences of DynTex [25] are used for building the three databases used in our experiments. The list of used image sequences for these databases are visible on the DynTex website[2]. A sample for each class of each database is show in Figure 7.

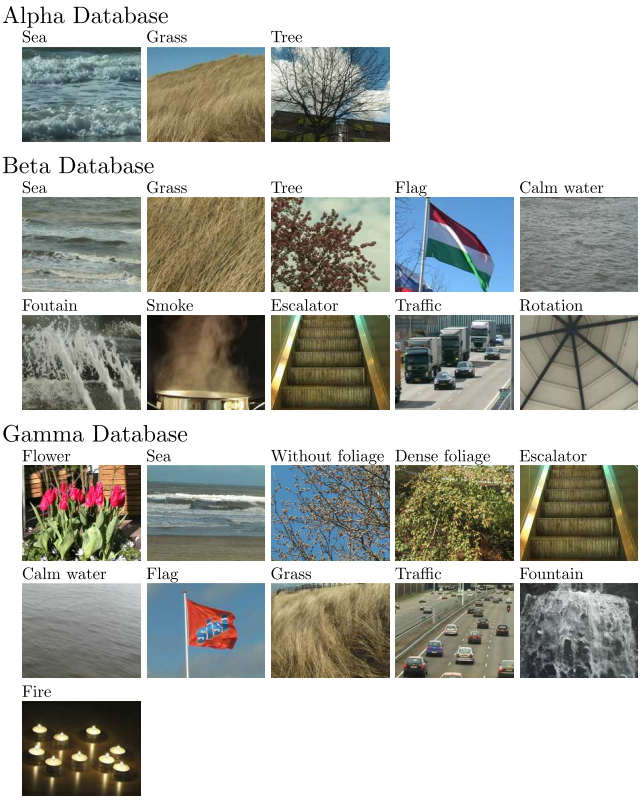

Fig. 7 *Sample for each class of each database.*

These databases differ in their complexity, their number of classes and their number of elements:

– **Alpha Database**: 60 image sequences of dynamic textures grouped in 3 relatively simple classes: sea, grass and trees.

[2] `http://projects.cwi.nl/dyntex/`

	c_1	c_2	c_3
c_1 - Sea [20]	**20**		
c_2 - Grass [20]	4	**15**	1
c_3 - Trees [20]		2	**18**

Table 2 Confusion matrix for the Alpha database when using as feature vector the coefficient of Spatial Wavelet Decomposition frame by frame.

- **Beta Database**: 162 image sequences of dynamic textures grouped in 10 classes: sea, grass, trees, flags, calm water, fountains, smoke, escalator, traffic, rotation. More complex phenomena are present here than in the Alpha Database.
- **Gamma Database**: 275 image sequences of dynamic textures grouped in 11 classes: flowers, sea, trees without foliage, dense foliage, escalator, calm water, flags, grass, traffic, fountains, fire. In this database, some classes are composed of many samples covering many cases (change in scale, orientation, etc.). This dataset is complex and challenging.

These three benchmarks are another contribution of this article. Indeed, previous experiments were conducted on unavailable or/and small databases.

3.3 Results

In this section, results of dynamic texture recognition are presented. All multiscale approaches are computed using five resolution levels. Only the method of J.R. Smith *et al.* Wavelet Decomposition [29] is performed on three levels, as its construction does not allow more decomposition levels.

Table 1 shows the obtained recognition rates with the feature vector previously exposed. Tables 2, 3 and 4 show the confusion matrices obtained with feature vector $\mathbf{S}_m^A$ respectively for Spatial Wavelet Decomposition on Alpha database, Morphological Component Analysis on Beta database and 2D+T Wavelet Decomposition on Gamma database. These confusion matrices represent the best recognition rates for each database.

These results lead to the following observations for the feature vectors:

- whatever the analysis methods of image sequences and databases used, the feature vectors S_m^μ is the least discriminating. Its association with other descriptors for creating S_m^A is questionable. Experiments where the vector S_m^A is built without S_m^μ were carried out and the obtained results do not have a better recognition rate, and in some cases it is even degraded.
- observations of only four feature vectors (S_m^μ, S_m^σ, S_m^E and S_m^H) show that feature vectors built with standard deviation of detail subbands have the best discriminative power.
- the association of different feature vectors for building S_m^A are beneficial. In most cases, the recognition rates using these features are better.

The observation of recognition rates for each database induces the following remarks:

- **Alpha database**: the most discriminant method is the Spatial Wavelet Decomposition applied frame per frame. The observation of the Alpha Database samples shows that the distinction between classes can be performed only with spatial properties. In this case, the temporal information does not add relevant information.
  The other analysis methods (except for Temporal Wavelet Decomposition) remain effective as they get a similar recognition rate (5% of difference).
- **Beta database**: for this database, the most discriminant method is based on Morphological Component Analysis. This one has the best recognition rate for four proposed feature vectors (S_m^σ, S_m^E, S_m^H and S_m^A).
- **Gamma database**: the 2D+T Wavelet Decomposition is the approach that obtained the best recognition rate for this database. For the other databases, recognition rates are similar than the other approaches.

The different multiscale approaches, except the temporal wavelet decomposition, give acceptable recognition rates for all databases. The obtained results are always nearby, while they are inferior if only temporal information is used.

The different multiscale approaches have been computed with different parameters (different decomposition levels, different thresholding strategies for Morphological Component Analysis, ...). Results shown in this section use the best parameter set.

3.4 Discussion and prospects

The obtained recognition rates are satisfactory: they are close to 70% for databases of substantial size (Beta and Gamma databases) using frequential information only and without any color information. These performances can be improved.

Table 5 gives the size of feature descriptors depending on the decomposition method. For some methods this size is greater than the number of samples. Difficulties met are the following:

Analysis Method	Database	$\mathbf{S}_m^{\mu}$	$\mathbf{S}_m^{\sigma}$	$\mathbf{S}_m^{E}$	$\mathbf{S}_m^{H}$	$\mathbf{S}_m^{A}$
Spatial Wavelet Decomposition frame per frame	Alpha	68	82	78	88 †⋆	88 †⋆
	Beta	41 †	50	51	57	66
	Gamma	38	60	60	55	65
Temporal Wavelet Decomposition	Alpha	37	75	67	67	73
	Beta	15	43	37	28	46
	Gamma	15	34	31	26	40
2D+T Wavelet Decomposition	Alpha	72 †	85 †	85 †	87	85
	Beta	33	62	61	65	65
	Gamma	36	65 †	64 †	61 †	68 †⋆
J.R. Smith *et al.* Wavelet Decomposition [29]	Alpha	65	83	80	82	83
	Beta	35	65	59	65	67
	Gamma	43 †	63	56	59	65
2D+T Curvelet Decomposition	Alpha	47	85 †	83	85	85
	Beta	19	65	61	62	67
	Gamma	18	62	60	56	63
Morphological Component Analysis	Alpha	37	83	83	83	85
	Beta	23	68 †	64 †	66 †	70 †⋆
	Gamma	19	61	62	59	63

Table 1 *Dynamic texture recognition rates (in %) for three databases according to the different computed feature vectors.* † *represents the best recognition rates for one feature vector in one database.* ⋆ *represents the best recognition rate of one database.*

	c_1	c_2	c_3	c_4	c_5	c_6	c_7	c_8	c_9	c_{10}
c_1 - Sea [20]	**19**						1			
c_2 - Vegetation [20]	1	**15**	1			3				
c_3 - Trees [20]		4	**15**			1				
c_4 - Flags [20]		1		**13**		3	3			
c_5 - Calm water [20]	3				**14**		3			
c_6 - Fountains [20]		6		1	1	**12**				
c_7 - Smoke [16]							**16**			
c_8 - Escalator [7]		1		2				**4**		
c_9 - Traffic [9]		1			2		1		**5**	
c_{10} - Rotation [10]		3		4		1			2	**0**

Table 3 Confusion matrix for the Beta database when using as feature vector the coefficient of Morphological Component Analysis.

	c_1	c_2	c_3	c_4	c_5	c_6	c_7	c_8	c_9	c_{10}	c_{11}
c_1 - Flowers [29]	**23**		1	4				1			
c_2 - Sea [38]		**35**				1					2
c_3 - Trees without foliage [25]	5		**15**	4				1			
c_4 - Dense foliage [35]	6		8	**18**				1	1	1	
c_5 - Escalator [7]					**4**		2	1			
c_6 - Calm water [30]		4				**19**	1		1	5	
c_7 - Flag [31]			1	1			**20**	2		1	6
c_8 - Grass [23]		3		2	2	2		**12**		1	1
c_9 - Traffic [9]	1								**7**		1
c_{10} - Fountains [37]	3					3		5		**25**	1
c_{11} - Fire [11]							1				**10**

Table 4 Confusion matrix for the Gamma database when using as feature vector the coefficient of the 2D+T Wavelet Decomposition.

– the information to classify samples is too redundant and can degrade the classification.
– comparing between multiscale methods can be discussed as it is not performed in the same experimental conditions. Indeed, the classification of 10 classes in a 3 dimensional space is not as difficult than in a 5508 dimensional space. Moreover, the used classification methods may not be adapted for high dimensional spaces.

For reducing the size of feature vectors, some approaches are proposed:

– reduce the feature dimension using Principal Component Analysis (PCA). An experiment has been carried out using PCA, where the dimension has been reduced to 50 for each possible feature vector (the size of feature vector are much larger than 50). The obtained recognition rates greatly decrease (-35% on average). The same observation can be done if we retain just 15 principal components instead of 50.
– to build a feature selection method, for example, the Stepwise Discriminant Analysis method [14]. Geometrically, it means finding the representation subspace that allows maximal distance between gravity centers of scattering. In the literature, many approaches enable to perform feature selection, for instance the recent work described in [34] that uses a sparse representation.
– to change the feature construction for obtaining a smaller and more representative spatio-temporal information. Rather than computing the energy of each detail subband for each multiscale approach, it is possible to compute only an energy at each scale and used other descriptors for characterizing directional information (as the directional homogeneity criterion [23]).
  For example, instead of having a vector of 3460 elements for the 2D+T Curvelet Decomposition, we get 5 elements representative of different scales plus directional information.
– for the Morphological Component Analysis, the feature vectors are built similarly for the two components. It could be relevant to build descriptors adapted to the extracted components.

4 Conclusion

This paper presents two approaches based on the 2D+T Curvelet Decomposition for characterizing dynamic textures. The first one use the geometrical multiscale decomposition and one the second is based on a decomposition of dynamic texture into two components (wavefront and local phenomena). Our goal is to study the influence of spatio-temporal decomposition on the dynamic texture characterization.

After presenting the different multiscale approaches for characterizing dynamic textures, we propose several features using detail subbands. These features are tested on three new large databases publicly available. Finally, results of dynamic textures recognition are presented and discussed.

Using multiscale approaches for analysing dynamic textures is very promising as it is closely linked to the physical properties of dynamic textures.

The descriptors used can be improved; for instance the different components obtained using the Morphological Component Analysis algorithm can be used for extracting characteristic features; some related to the geometry of the dynamic texture (main motion direction, uniformity of the global movement, etc.) and some characterizing more local phenomena (speed, local vortex, etc.). Our future works will be focused on the invariance to rotation and scale of descriptors.

With efficient dynamic texture descriptors, many other applications can be considered: tracking of dynamic texture (evolution of a fire), synthesis (realistic rendering of dynamic textures in video games and animation film), indexing, etc..

References

1. Ali, I., Mille, J., Tougne, L.: Space–time spectral model for object detection in dynamic textured background. Pattern Recognition Letters **33**(13), 1710–1716 (2012)
2. Aujol, J., Chambolle, A.: Dual Norms and Image Decomposition Models. Computer Vision **63**, 85–104 (2005)
3. Candès, E.: Ridgelets : Theory and Applications. Ph.D. thesis, University of Standford (1998)
4. Candès, E., Demanet, L.: The curvelet Representation of Wave Propagators is Optimally Sparse. Communications on Pure and Applied Mathematics **58**, 1472–1528 (2005)
5. Candès, E., Demanet, L., Donoho, D., Ying, L.: Fast Discrete Curvelet Transforms. Tech. rep., California Institute of Technology (2005)
6. Chan, A., Vasconcelos, N.: Modeling, Clustering, and Segmenting Video with Mixtures of Dynamic Textures. IEEE Transactions on Pattern Analysis and Machine Intelligence **30**, 909–926 (2008)
7. Chan, A.B., Mahadevan, V., Vasconcelos, N.: Generalized Stauffer-Grimson background subtraction for dynamic scenes. Machine Vision and Application **22**(5), 751–766 (2011)
8. Chetverikov, D., Peteri, R.: A Brief Survey of Dynamic Texture Description And Recognition. In: International Conference Computer Recognition Systems, pp. 17–26. Springer (2005)
9. Donoho, D., Duncan, M.: Digital Curvelet Transform: Strategy, Implementation and Experiments. In: Wavelet Applications VII, pp. 12–29. SPIE (1999)
10. Doretto, G., Chiuso, A., Wu, Y., Soatto, S.: Dynamic Textures. International Journal of Computer Vision **51**, 91–109 (2003)

Analysis method	Number of scales	Number of subdivisions	Size of feature vectors	
			S_m^d $d = \{\mu, \sigma, E, H\}$	S_m^A
Spatial Wavelet Decomposition frame per frame	3	×	9	36
	4	×	12	48
	5	×	15	60
Temporal Wavelet Decomposition	3	×	3	12
	4	×	4	16
	5	×	5	20
2D+T Wavelet Decomposition	3	×	21	84
	4	×	28	112
	5	×	35	140
J.R. Smith Wavelet Decomposition	3	×	45	180
2D+T Curvelet Decomposition	3	2	25	100
	3	4	97	388
	4	2	121	484
	4	4	481	1924
	5	2	217	868
	5	4	865	3460
Morphological Component Analysis	×	×	1377	5508

Table 5 *Size of feature vectors depending on the analysis methods.*

11. Dubois, S., Péteri, R., Ménard, M.: A 3D Discrete Curvelet based Method for Segmenting Dynamic Textures. In: International Conference on Image Processing (ICIP 09), pp. 1373–1376 (2009)
12. Dubois, S., Péteri, R., Ménard, M.: A Comparison of Wavelet Based Spatio-temporal Decomposition Methods for Dynamic Texture Recognition. In: Iberian Conference on Pattern Recognition and Image Analysis (IbPRIA 09), pp. 314–321 (2009)
13. Dubois, S., Péteri, R., Ménard, M.: Decomposition of Dynamic Textures using Morphological Component Analysis. IEEE Transactions on Circuits and Systems for Video Technology **22**(2), 188–201 (2012)
14. Duda, R.O., Hart, P.E., Stork, D.G.: Pattern Classification (2nd Edition). Wiley-Interscience (2001)
15. Fablet, R., Bouthemy, P.: Motion Recognition Using Spatio-temporal Random Walks in Sequence of 2D Motion-related Measurements. In: IEEE International Conference on Image Processing (ICIP 01), pp. 652–655 (2001)
16. Finch, M.: GPU Gems: Programming Techniques, Tips, and Tricks for Real-Time Graphics, Chap.1. Randima Fernando (2004). DOI http://http.developer.nvidia.com/GPUGems/gpugems_part01.html. URL `http://http.developer.nvidia.com/GPUGems/gpugems_part01.html`
17. Li, J., Chen, L., Cai, Y.: Dynamic Texture Segmentation Using 3-D Fourier Transform. In: International Conference on Image and Graphics (ICIG 09), pp. 293–298 (2009)
18. Mallat, S., Peyré, G.: A Review of Bandlet Methods for Geometrical Image Representation. Numerical Algorithms **44**(3), 205–234 (2007)
19. Meyer, Y.: Workshop sur "an interdisciplinary approach to textures and natural images processing". Institut Henri Poincaré (2007)
20. Nelson, R., Polana, R.: Qualitative Recognition of Motion using Temporal Texture. Computer Vision and Image Understanding **56**, 78–89 (1992)
21. Otsuka, K., Horikoshi, T., Suzuki, S., Fujii, M.: Feature Extraction of Temporal Texture Based on Spatiotemporal Motion Trajectory. In: International Conference on Pattern Recognition (ICPR 98), p. 1047 (1998)
22. Péteri, R.: Tracking Dynamic Textures using a Particle Filter Driven by Intrinsic Motion Information. Machine Vision and Applications pp. 1–9 (2010)
23. Péteri, R., Chetverikov, D.: Qualitative Characterization of Dynamic Textures for Video Retrieval. In: International Conference on Computer Vision and Graphics (ICCVG 04), pp. 33–38 (2004)
24. Péteri, R., Chetverikov, D.: Dynamic Texture Recognition Using Normal Flow and Texture Regularity. In: Iberian Conference on Pattern Recognition and Image Analysis (IbPRIA 05), pp. 223–230 (2005)
25. Péteri, R., Fazekas, S., Huiskes, M.: DynTex: A Comprehensive Database of Dynamic Textures. Pattern Recognition Letters **31**, 1627–1632 (2010)
26. Phillips, W., Shah, M., Lobo, N.: Flame Recognition in Video. Pattern Recognition Letters **23**, 319–327 (2002)
27. Polana, R., Nelson, R.: Recognition of Motion from Temporal Texture. In: Conference on Computer Vision and Pattern Recognition (CVPR 92) (1992)
28. Saisan, P., Doretto, G., Wu, Y., Soatto, S.: Dynamic Texture Recognition. In: Conference on Computer Vision and Pattern Recognition (CVPR 01), pp. 58–63 (2001)
29. Smith, J., Lin, C., Naphade, M.: Video Texture Indexing using Spatio-Temporal Wavelets. In: IEEE International Conference on Image Processing (ICIP 02), pp. 437–440 (2002)
30. Starck, J., Elad, M., Donoho, D.: Redundant Multiscale Transforms and their Application for Morphological Component Analysis. Advances in Imaging and Electron Physics **132** (2004)
31. Wildes, R.P., Bergen, J.R.: Qualitative Spatiotemporal Analysis Using an Oriented Energy Representation. In: European Conference on Computer Vision (ECCV00), pp. 768–784 (2000)
32. Witten, I.H., Frank, E.: Data Mining: Practical Machine Learning Tools and Techniques, Second Edition (Morgan Kaufmann Series in Data Management Systems). Morgan Kaufmann Publishers Inc., San Francisco, CA, USA (2005)
33. Wu, P., Ro, Y., Won, C., Choi, Y.: Texture Descriptors in MPEG-7. In: International Conference on Computer Analysis of Images and Patterns (CAIP 01), pp. 21–28 (2001)
34. Xiang, S., Nie, F., Meng, G., Pan, C., Zhang, C.: Discriminative least squares regression for multiclass classification and feature selection. Neural Networks and Learning Systems, IEEE Transactions on **23**(11), 1738–1754 (2012)

35. Xiang, S., Nie, F., Zhang, C.: Learning a mahalanobis distance metric for data clustering and classification. Pattern Recognition **41**(12), 3600 – 3612 (2008)
36. Ying, L., Demanet, L., Candès, E.: 3D Discrete Curvelet Transform. In: International Society for Optical Engineering (SPIE 05) (2005)
37. Zhao, G., Pietikäinen, M.: Dynamic Texture Recognition Using Local Binary Patterns with an Application to facial expressions. IEEE Transactions on Pattern Analysis and Machine Intelligence **29**, 915–928 (2007)
38. Zhong, J., Scarlaroff, S.: Temporal Texture Recognition Model Using 3D Features. Tech. rep. (2002)